

Credit EDA

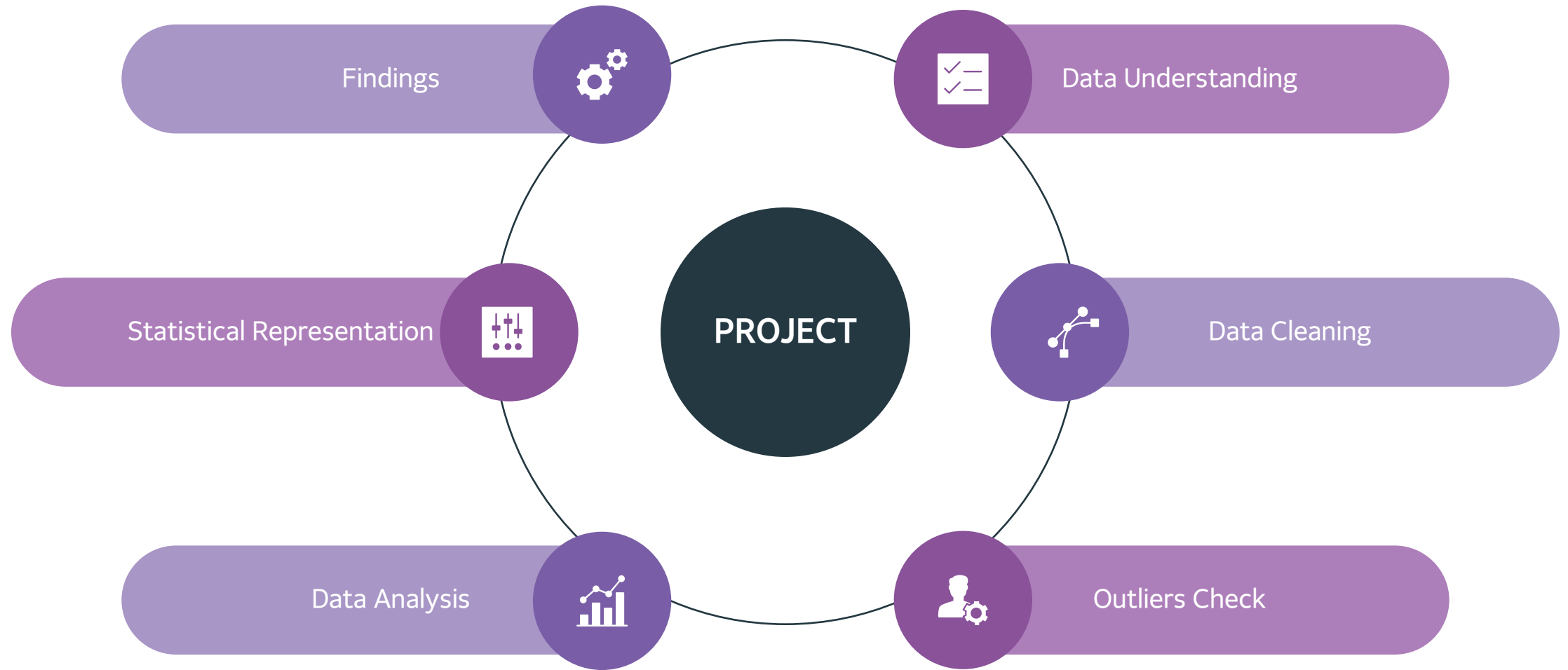
Anshul Khanna



PROBLEM STATEMENT

- To analyse data on loans and find pattern that predicts of client having difficulty of payment in the future.
- This has two risks associated while providing loans:
 - If the loan is not provided to the customers who can pay-back then it will be loss for the company.
 - If the loan is provided to the customers who cannot pay-back then it will also be loss for the company.

Credit EDA



DATA UNDERSTANDING

- Importing Libraries
- Loading the data
- Checking the structure of the data-frames
- Checking the data types, null values

DATA CLEANING

- Data quality and missing values validation
- Dropping columns with more than 40% of missing values.
- Imputing the columns with mean/median/mode.
- Changing datatypes to date, time, string, int, bool, etc for ease of analysis.

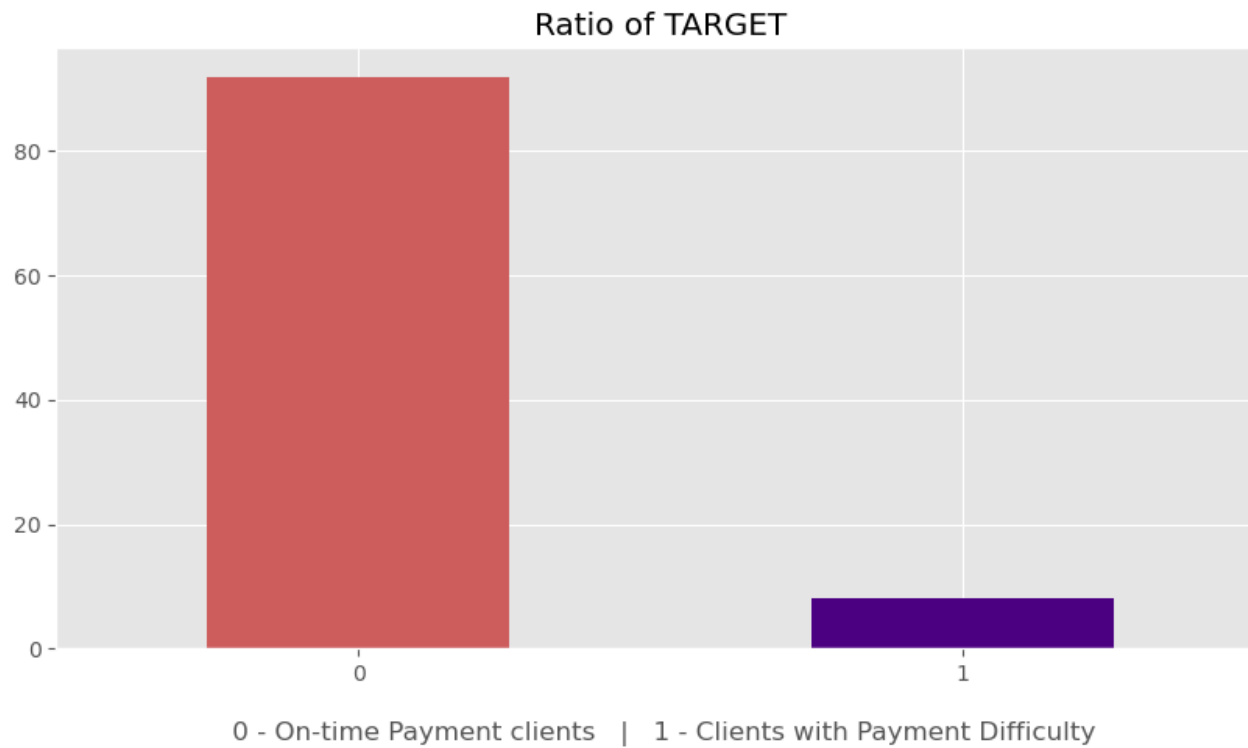
OUTLIERS

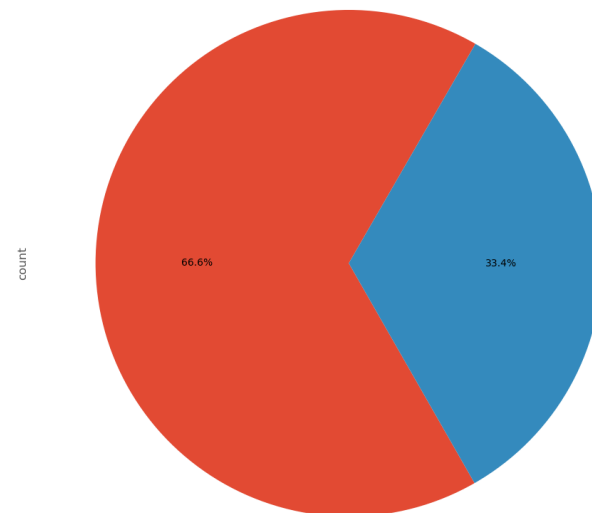
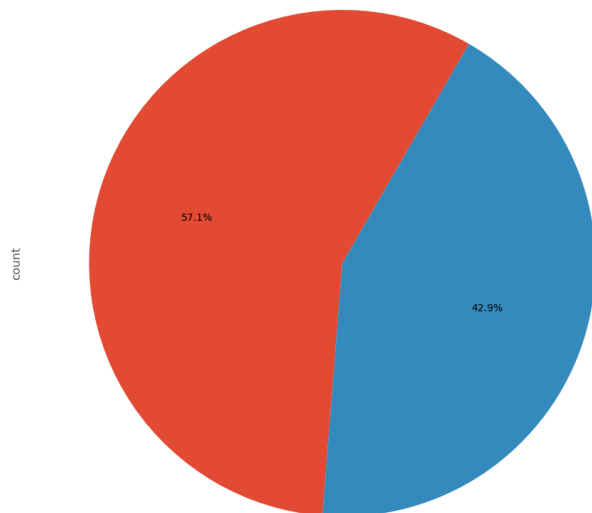
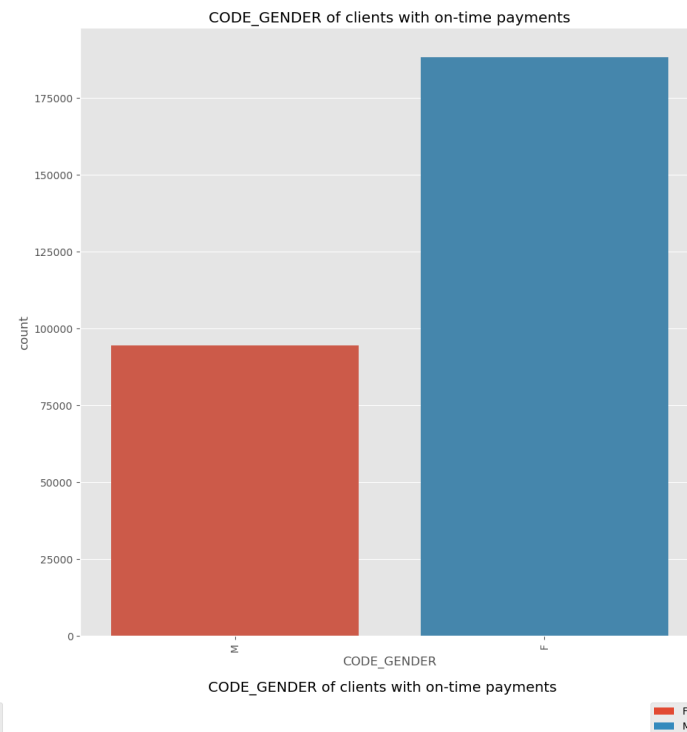
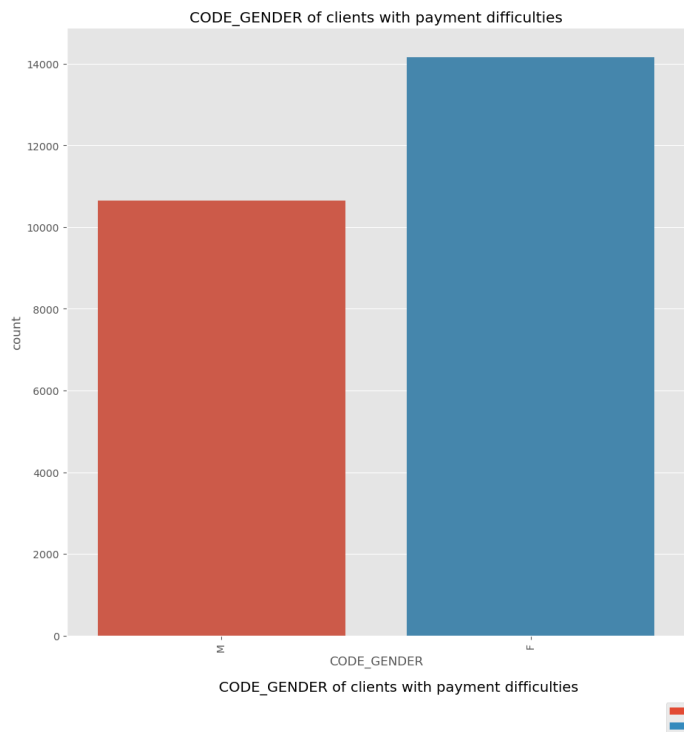
- Checking for outliers in the data, if they can create any bias in the analysis.
- Checking for imbalances and imbalance ratios.

DATA ANALYSIS

TARGET

- 1 in every 11 clients has payment difficulties

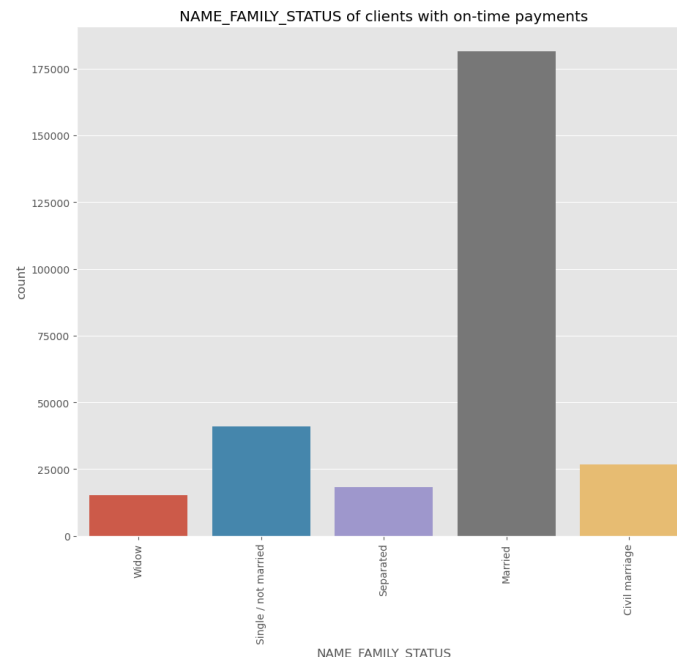
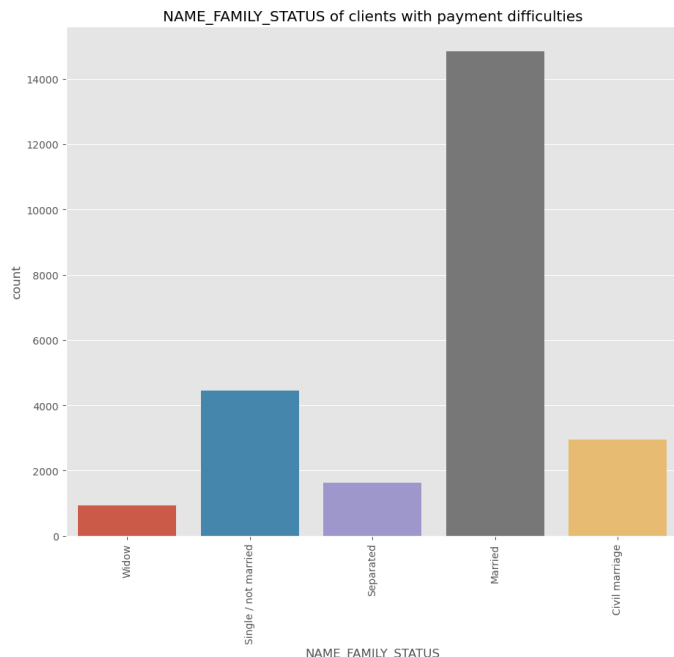




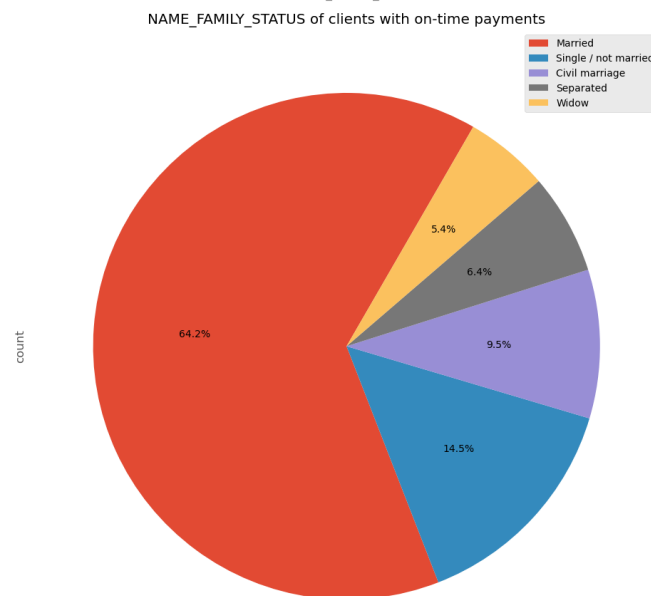
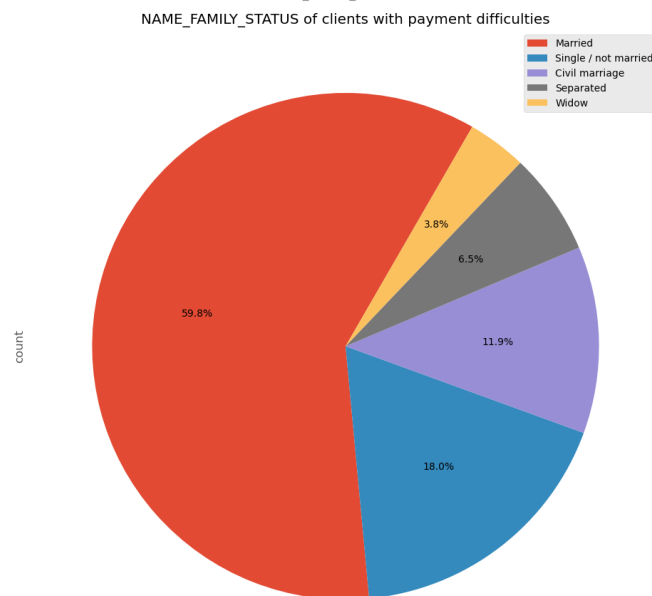
GENDER

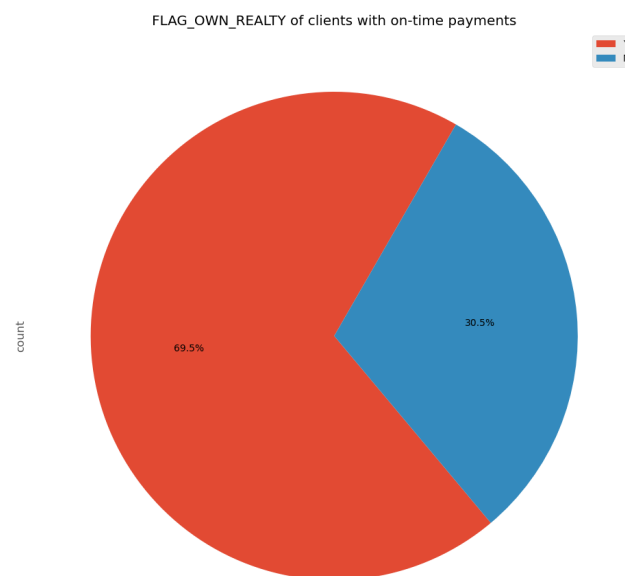
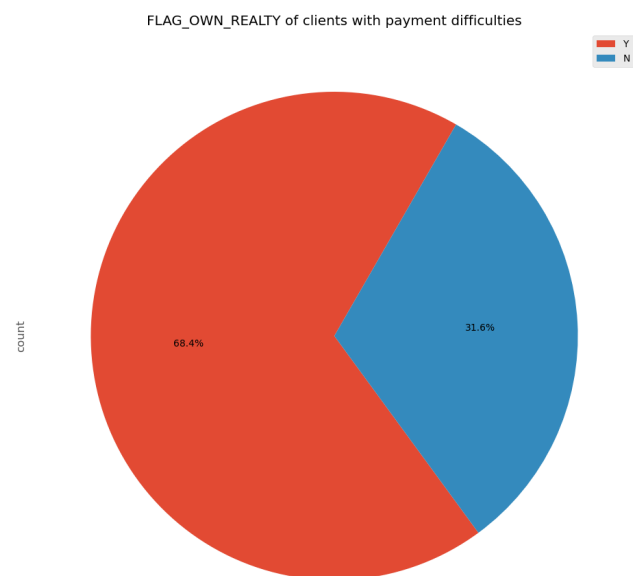
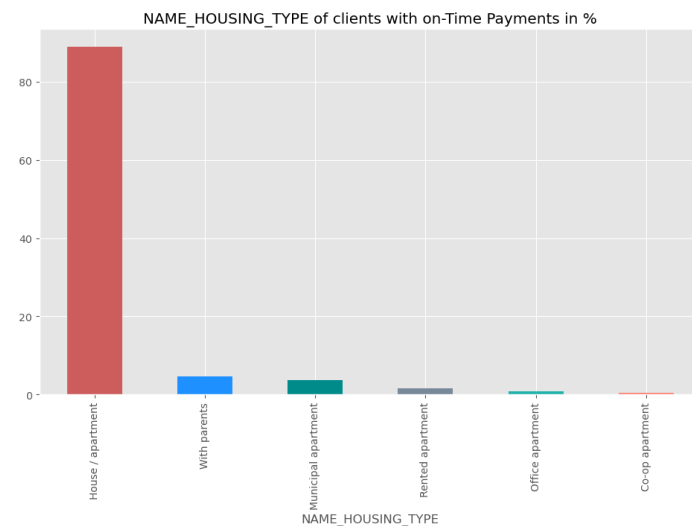
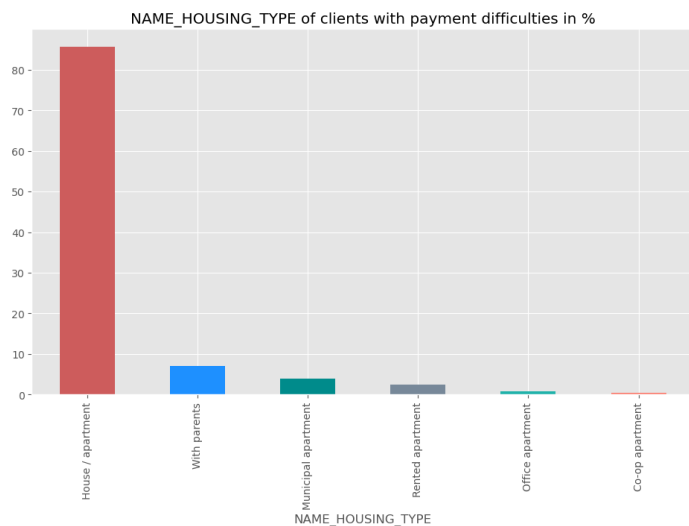
- Female clients are more than males.
- For clients with difficulties in payment: 57.1% is female and 42.9% is male.
- For clients with on time payment: 66.6% is female and 33.4% is male.

MARITAL STATUS



- Clients with married status are major in number.
- Married or Widowed are making more on time payments comparatively.
- Single/Non-Married have more payment difficulties.



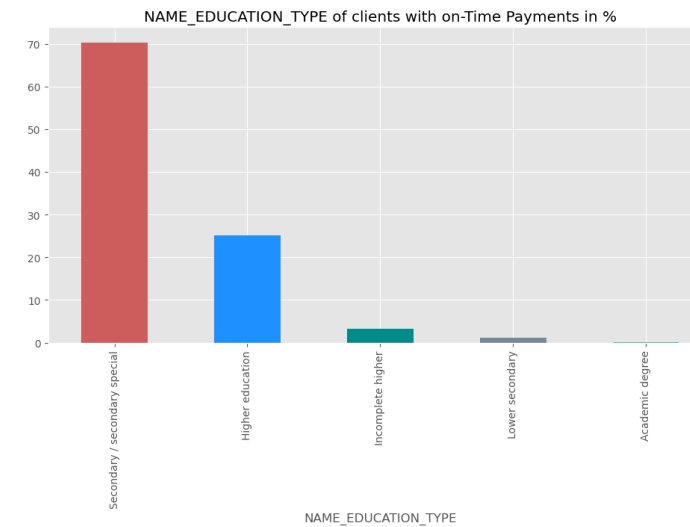
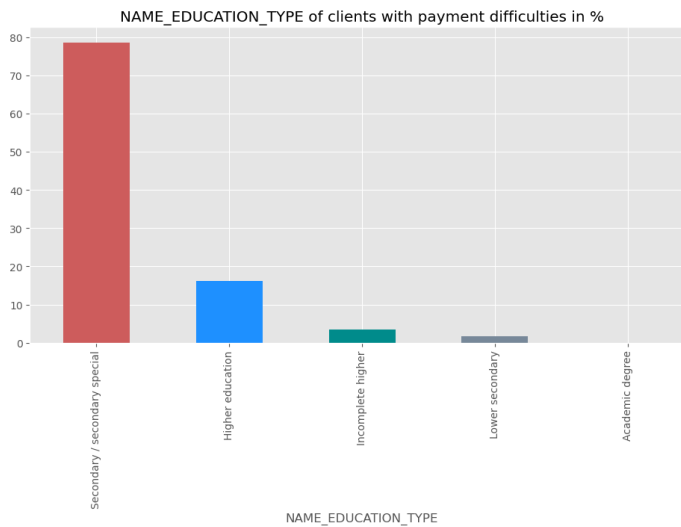
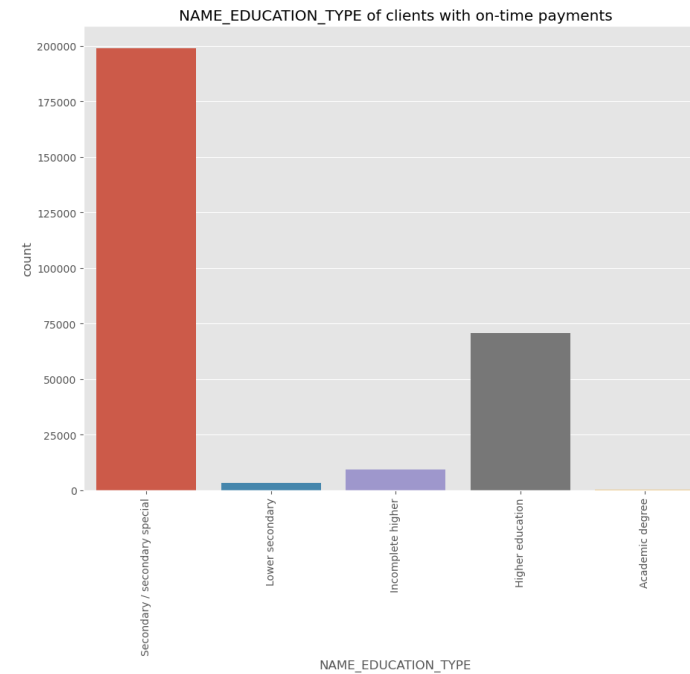
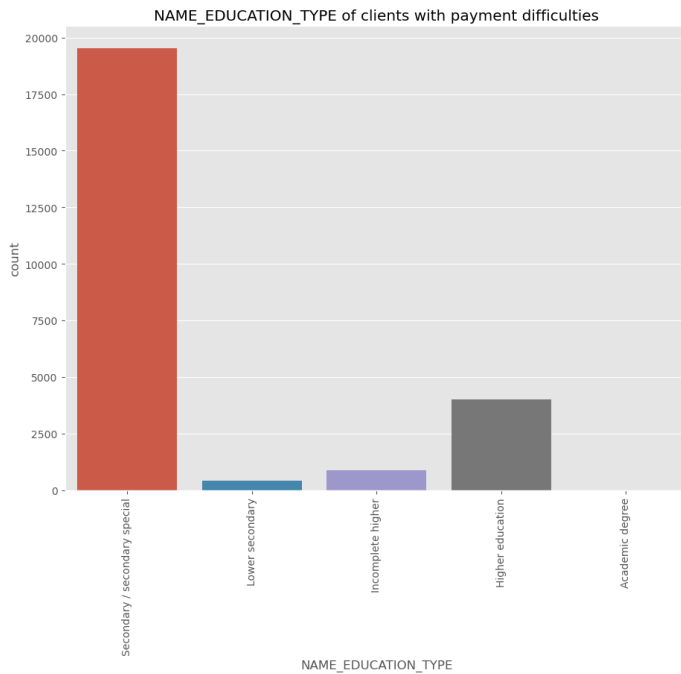


HOUSING

- The clients with home-owners are more in comparison to non-home-owners.
- For clients with difficulties in payment: 68.4% are home-owners and 31.6% are non-home-owners.
- For clients with on time payment: 69.5% are home-owners and 30.5% are non-home-owners.

EDUCATION

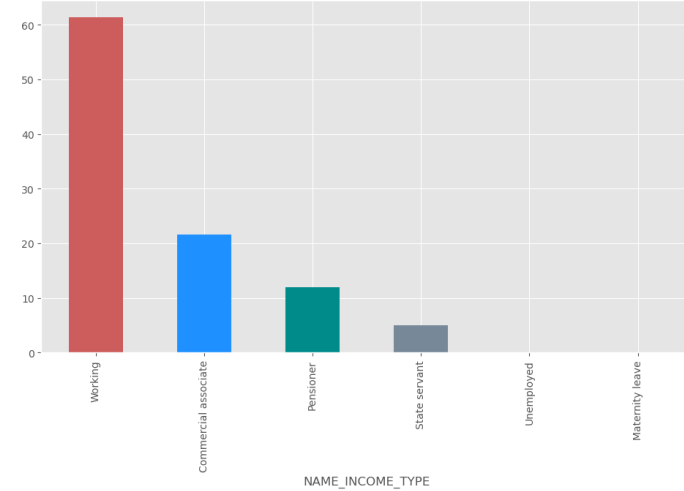
- Clients with higher education are paying on time.



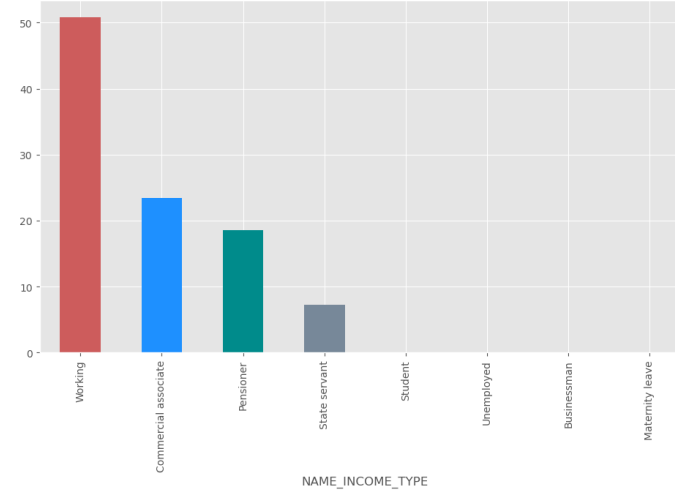
JOB TYPE

- Pensioners and state servants are more likely to make on time payments than delay the payments.

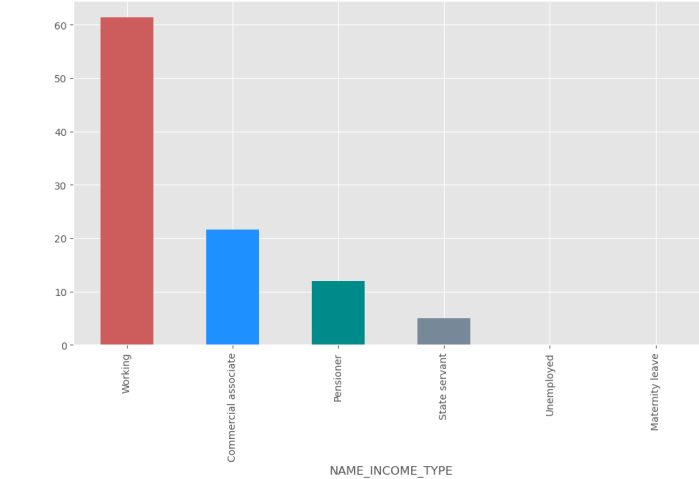
NAME_INCOME_TYPE of clients with payment difficulties in %



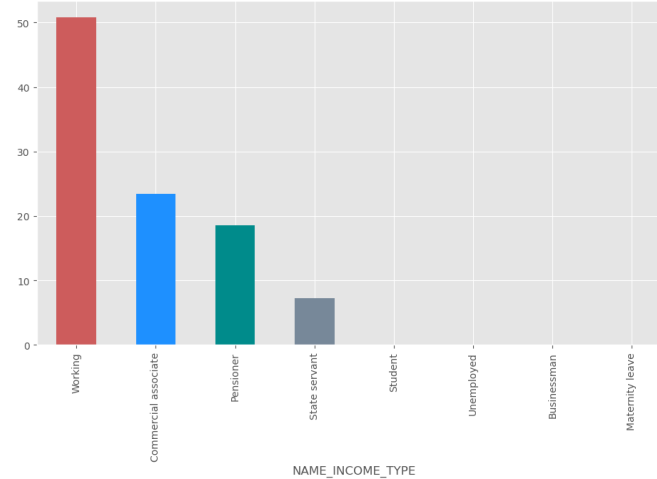
NAME_INCOME_TYPE of clients with on-Time Payments in %

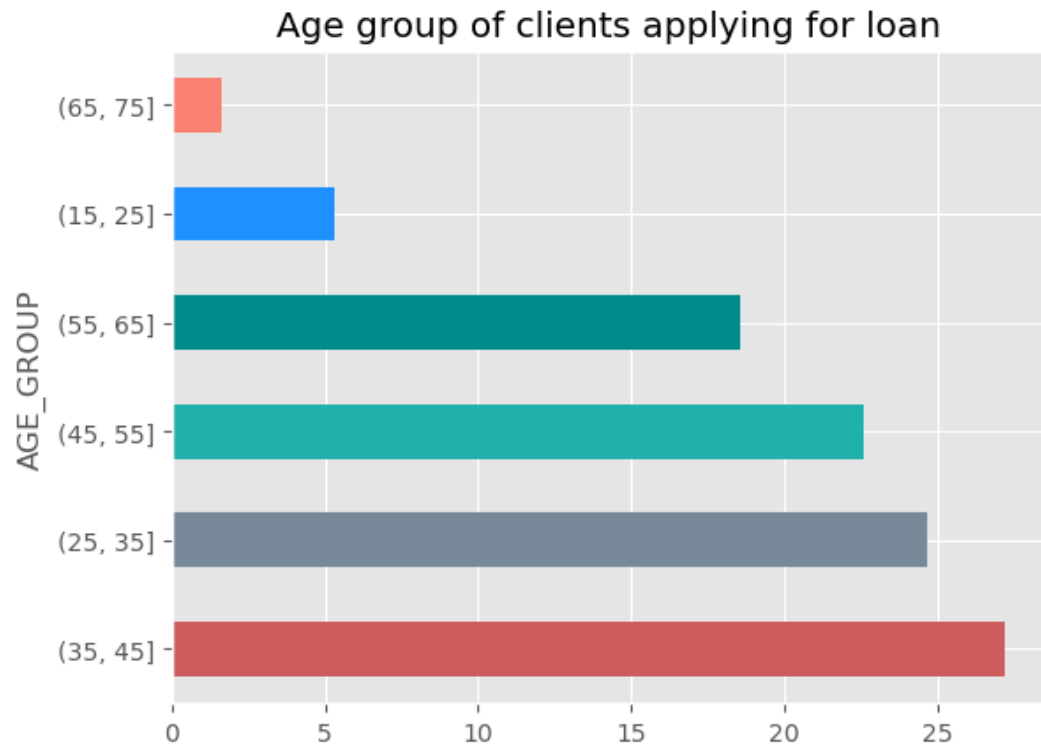
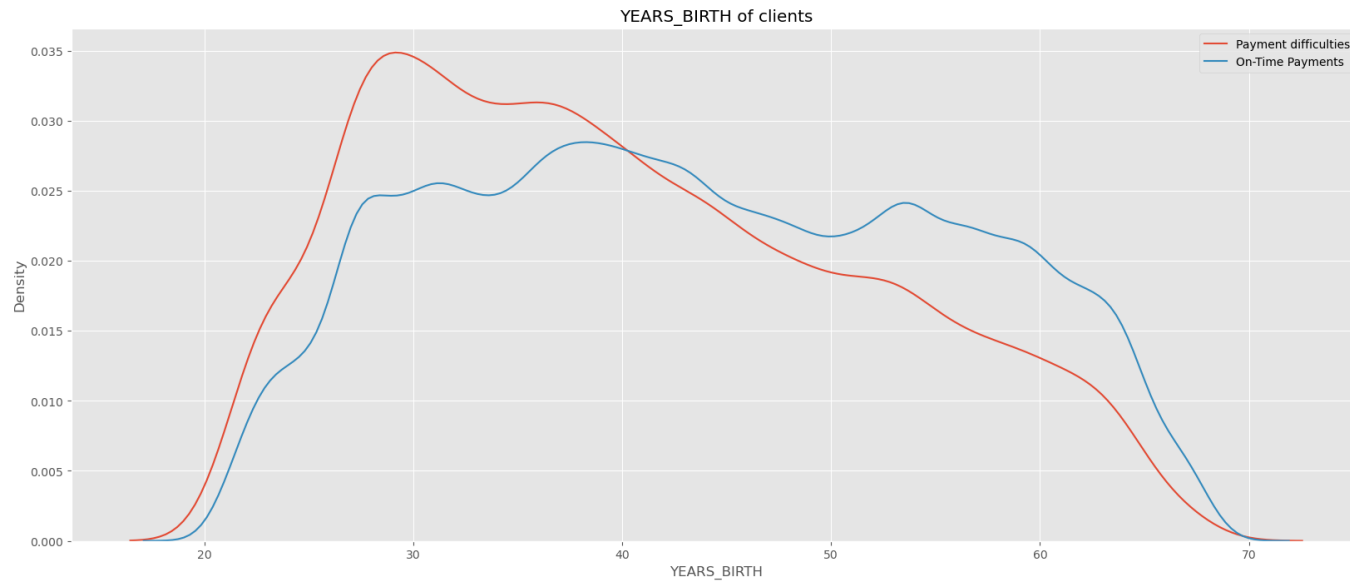


NAME_INCOME_TYPE of clients with payment difficulties in %



NAME_INCOME_TYPE of clients with on-Time Payments in %

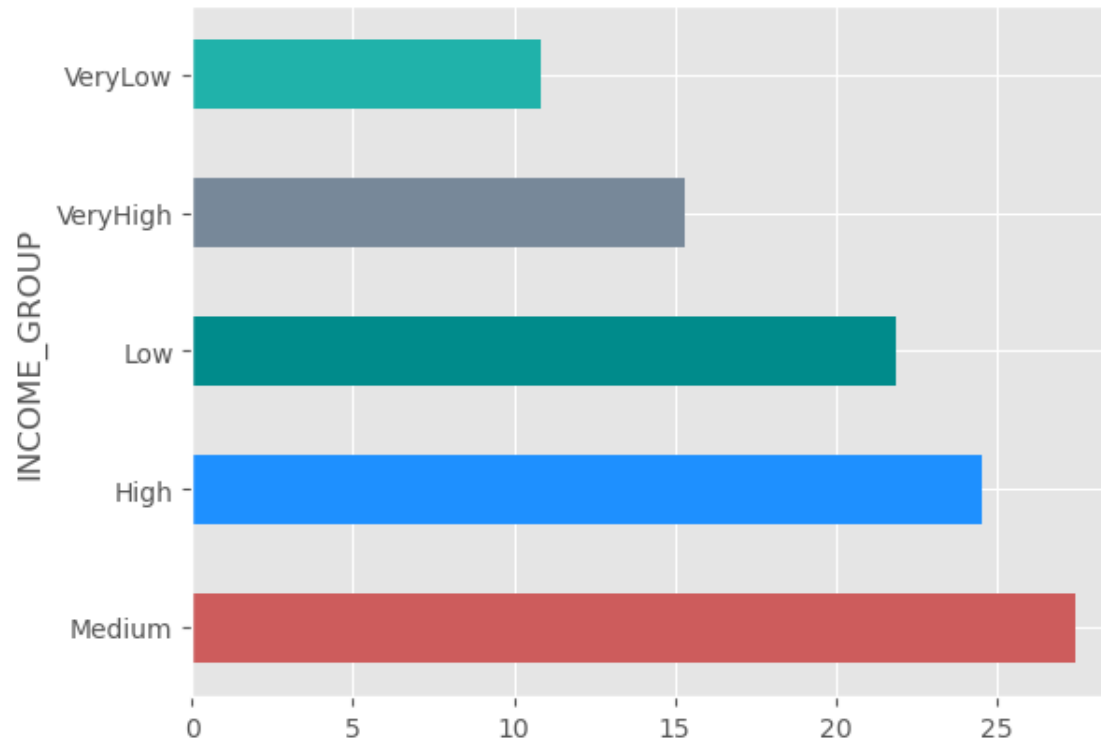




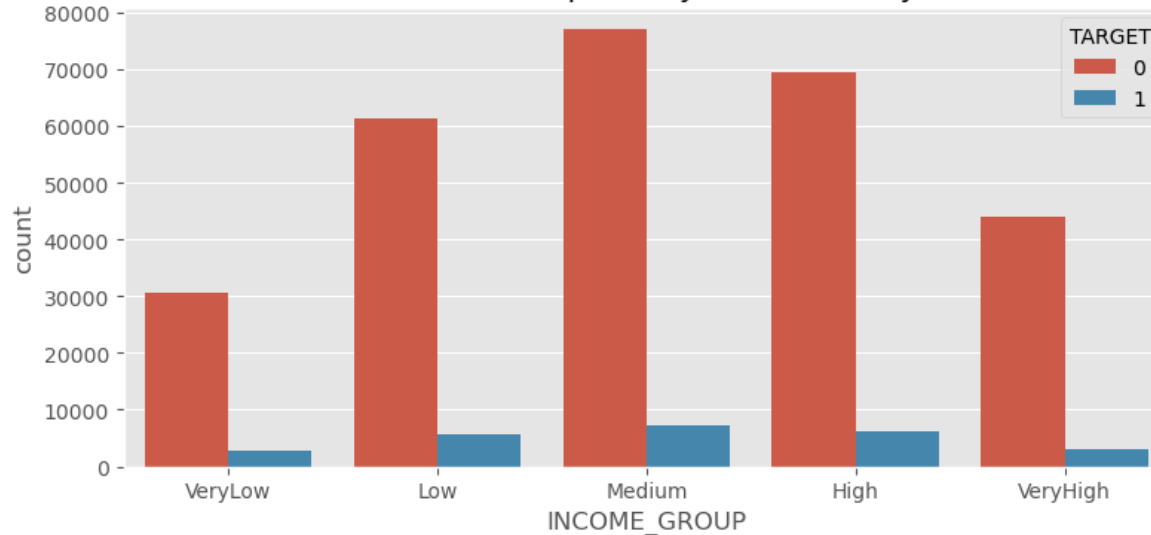
AGE

- Clients of age 20-40 years have more payment difficulties.
- Clients of age greater than 40 pay on time.
- Age 35-45 years are applying the most for loans.

Salary group of clients applying for loan

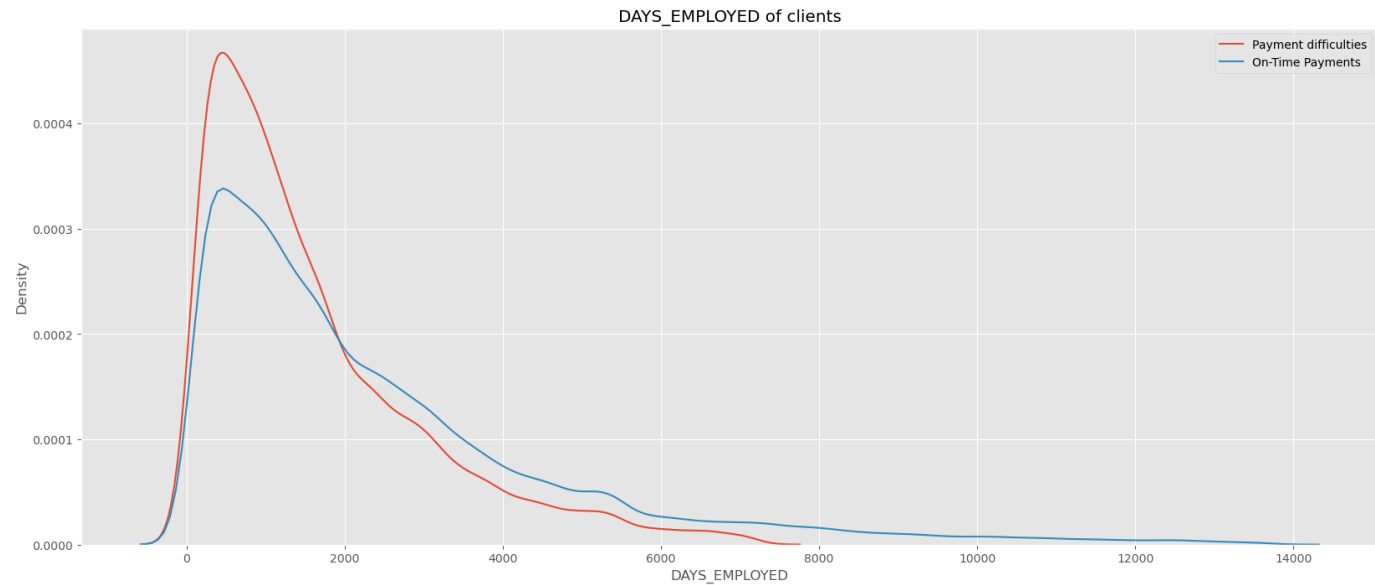


Income Group VS Payment Difficulty



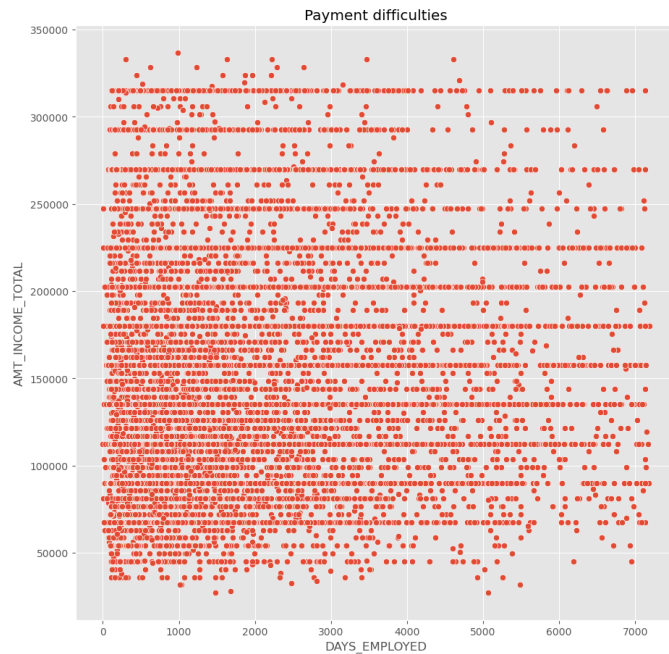
INCOME

- Medium income groups comprises of the greatest number of applicants.
- Default value per loan is highest in high income group.



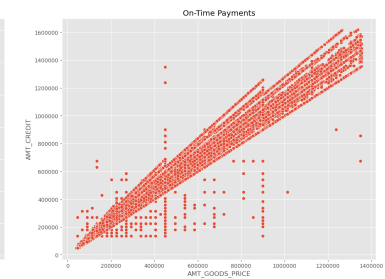
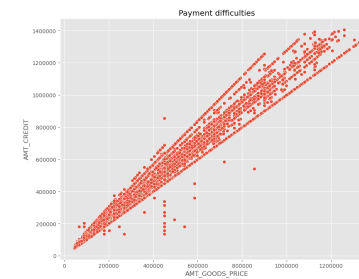
EMPLOYMENT DAYS

- For clients employed for less than 2000 days have payment difficulties.



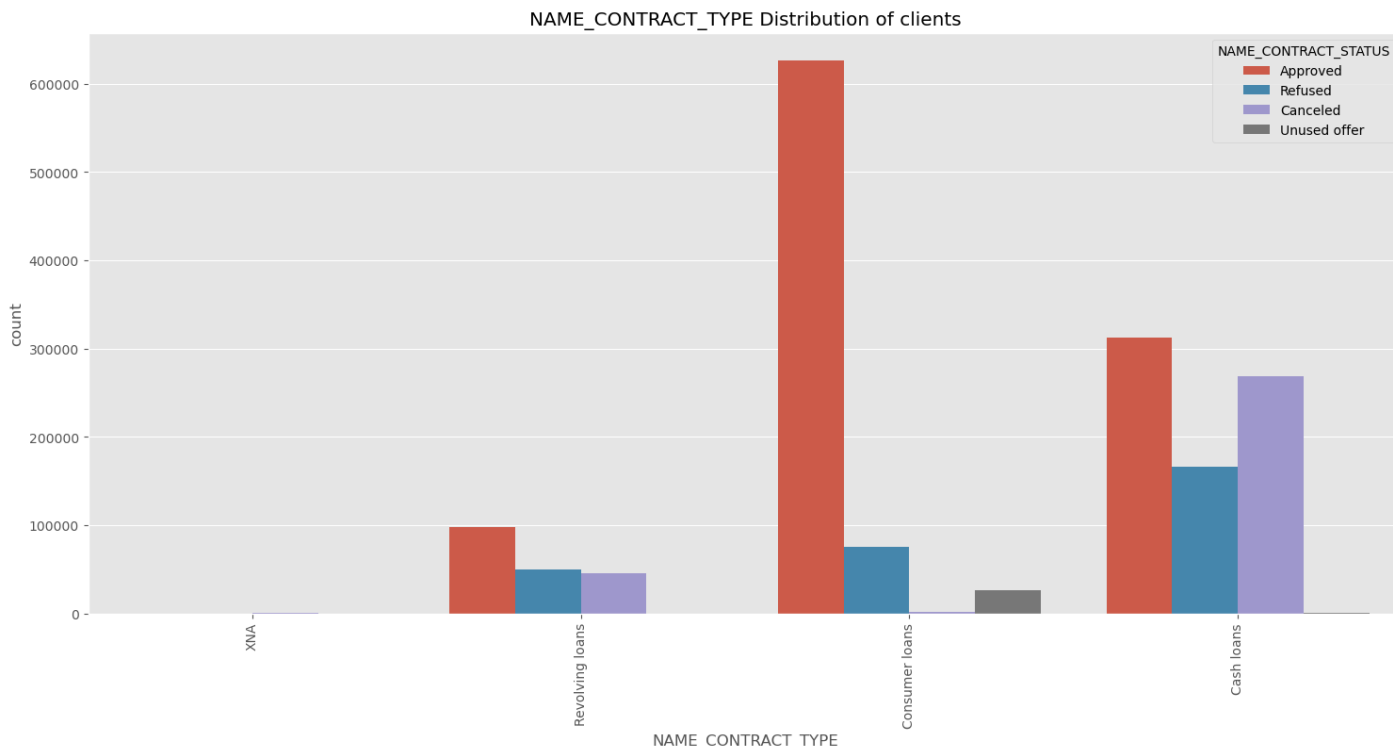
- The clients with more than 2000 days of employment pay on time.
- Clients with more than 7000 days of employment have very less number of payment difficulties.

CORRELATION



TOP CORRELATIONS

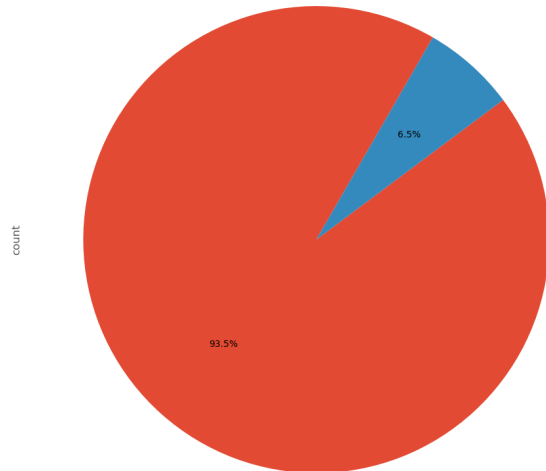
- AMT_CREDIT and AMT_GOODS_PRICE - 0.982
- REGION_RATING_CLIENT and REGION_RATING_CLIENT_W_CITY - 0.956637
- CNT_FAM_MEMBERS and CNT_CHILDREN - 0.885484
- DEF_60_CNT_SOCIAL_CIRCLE and DEF_30_CNT_SOCIAL_CIRCLE - 0.869016
- LIVE_REGION_NOT_WORK_REGION and REG_REGION_NOT_WORK_REGION - 0.847885
- REG_CITY_NOT_WORK_CITY and LIVE_CITY_NOT_WORK_CITY - 0.778540
- AMT_GOODS_PRICE and AMT_ANNUITY - 0.752295
- AMT_CREDIT and AMT_ANNUITY - 0.752195
- YEARS_BIRTH and YEARS_EMPLOYED - 0.582443



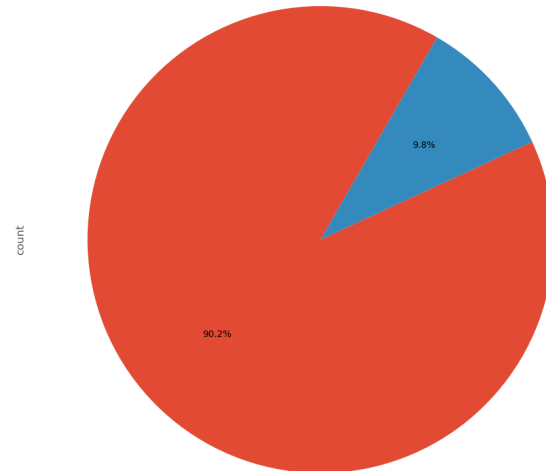
TYPES OF LOANS

- Consumer loans are approved most followed by cash loans.
- More cash loans have been refused than consumer loans.
- The bank has more repeaters in all approved, refused, unused, cancelled categories

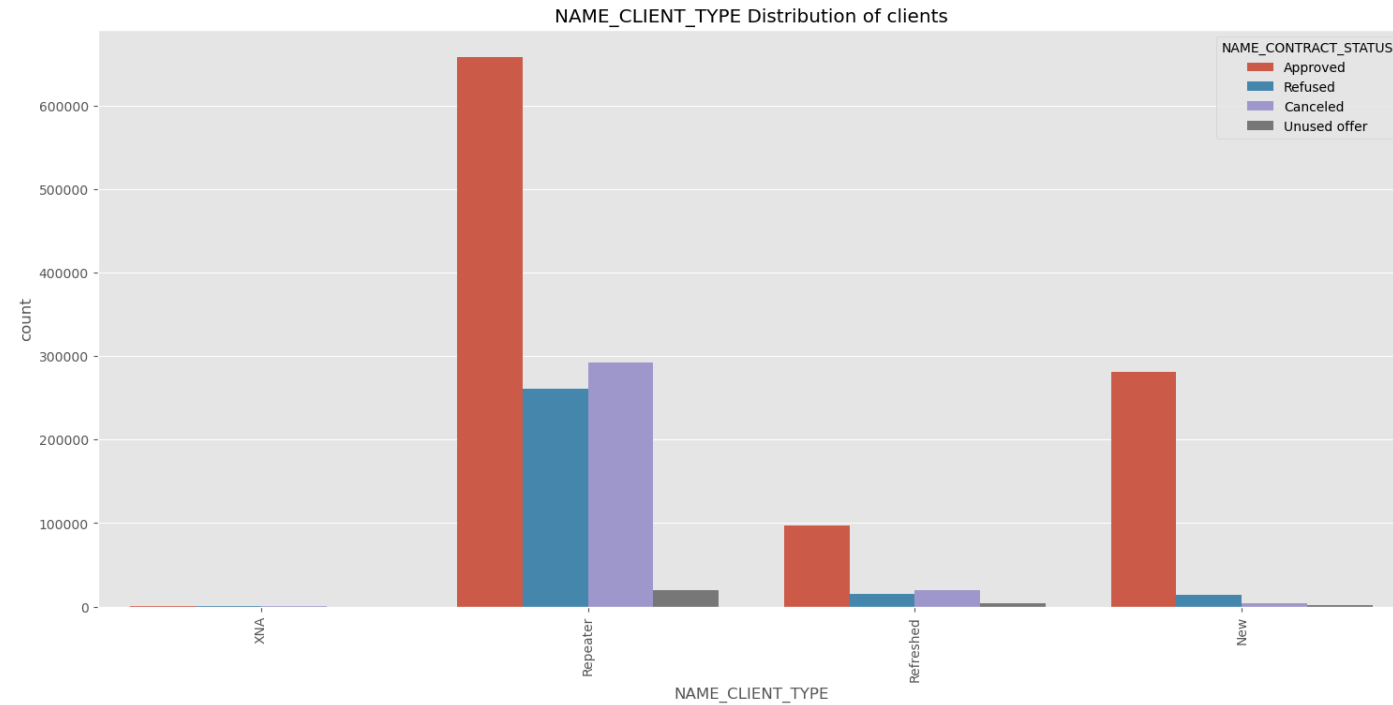
NAME_CONTRACT_TYPE of clients with payment difficulties



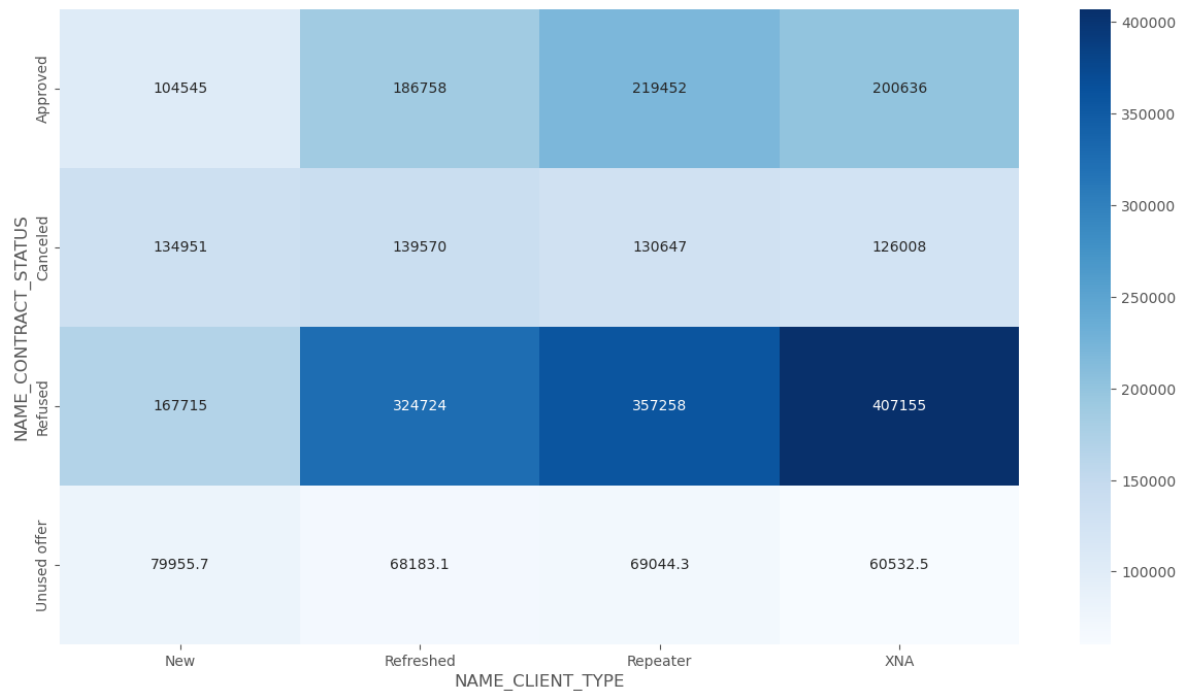
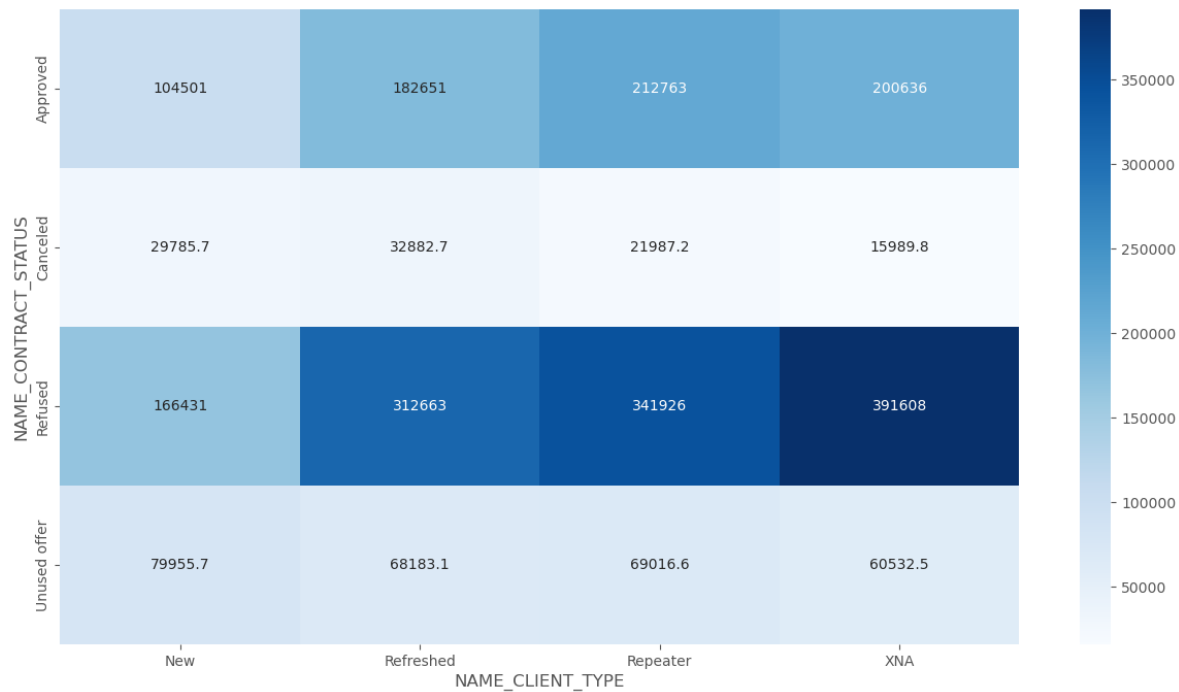
NAME_CONTRACT_TYPE of clients with on-time payments



CLIENT TYPES



- The number of repeat clients is more in all categories.
- Higher number of clients are cancelling a loan belong to repeaters.



REPEATERS CANCELING

- Refused cases have higher value amount.
- Low unused credit amount might be the reason for repeat customers to cancel.

SUMMARY

CHANCES OF CLIENT HAVING PAYMENT DIFFICULTY

- Based on Application Data-frame and checking against approved loans; the payment difficulties are observed in below:
 - Medium income
 - 25-35 years old, followed by 35-45 years age group
 - Male
 - Unemployed
 - Laborers, Salesman, Drivers
 - Not owing a house

OTHER FACTORS TO BE CONSIDERED

- No of Bureau Hits in last week. Month etc – zero hits is good
- Income 'Low' and 'High' is a concern
- Previous applications with Refused, Cancelled, Unused loans also have default which is a matter of concern.
- Credible Applications refused
- Unused applications have lower loan amount.
- Female applicants with payment difficulties are less.
- Students and Business mean have no problem in repayment of the loan on time.
- Previous applications with Refused, Cancelled, Unused loans also have cases where payments are coming on time in current application. This indicates that possibly wrong decisions were done in those cases.